Minimizing Cost Using the Tools and Services Provided by MGSS's AMMOS-PDS Pipeline Service (APPS)

Response to RFI: Preparation for the Development of a Community-Based Roadmap for NASA's Planetary Data Services
Solicitation Number: NNH15ZDA012L
Submitter: Stirling Algermissen (Stirling.S.Algermissen@jpl.nasa.gov)
Institution: Jet Propulsion Laboratory

Topic Addressed:

How can the interaction between the PDS and data providers (missions and individual researchers) be improved in order to make the archiving process seamless and less costly (to both data providers and the PDS)?

Rationale:

NASA missions have limited resources and often face tough decisions between funding hardware capability, ground system development, operations, and archive generation. In addition, high quality archive production has historically been expensive and time consuming. A significant portion of this cost involves iterative work between the mission and the PDS Discipline Node to define, document, peer review and validate the archive contents. By improving the mission-PDS interaction and reducing the cost and duration to generate the archive, scarce budgetary and schedule resources become available for other areas of the mission.

Suggested Improvements:

Often the effort to design, document and validate archives is repeated from one mission to the next, without the benefit of inherited tools or corporate knowledge transfer between missions.

On MSL many steps were taken to ensure high quality data sets were generated in a timely manner. Sharing MSL's lessons learned and how future Mars missions such as InSight and Mars 2020 intend to archive their data is beneficial to a wide range of planetary scientists.

Labels need to be designed and Software Interface Specification (SIS) documents must be written. Often times the data provider must learn how to meet PDS's required format. Once the archive has been designed the data provider must build it and correct any issues that come up during processing or later in a mission. It is not uncommon for a SIS document or label to change significantly throughout the lifetime of a mission. With the recent development of PDS4, data providers face more challenges than in the past to build an archive. PDS4's complex design and requirements are difficult for non-experts that do not have experience writing XML documents, schemas, and schematrons. There are numerous interactions between the PDS and the data provider where improvements can be made to make the archiving process seamless and less costly.

To decrease the cost and time to build an instrument archive MSL embedded a member of the PDS into MSL's OPGS (Operations Product Generation Subsystem) team. OPGS is responsible for processing MSL's data products and maintaining MSL's instrument SIS documents. Having a member of the PDS work closely with OPGS

significantly decreased the amount of time between an issue being discovered and resolved in an archive release. The PDS team member was able to resolve issues directly with the developers that wrote the processes that build the data products and metadata labels for the archive. OPGS and PDS also shared the machine that was used to build MSL's instrument volume. This allowed OPGS and the PDS to view the status of an archive as it was being built and deploy fixes quickly as needed. It also removed the need for OPGS and PDS to go back and forth using emails to resolve an issue. A close collaboration between a member of the PDS and the data provider made building a PDS archive much more seamless for the MSL. Even before archive building began it was found that having a dedicated point of contact early on was beneficial. A close working relationship also provides an avenue for finding and fixing problems early on, before large scale data production begins. Experience from MSL has shown the interaction between PDS and data providers can be improved by ensuring that the PDS and the data provider work closely together.

MSL discovered many issues in its data product labels during its first year of operations. For MSL's Sol 0-90 archive release #1 archive issues were not discovered until the manual archive building process began – months after the data products had been generated for operations. Had MSL known about these issues when they arose fixes could have been developed earlier and a higher quality data set could have been delivered to the PDS. As the mission progressed and new and different observations were recorded other issues were discovered relating to the metadata labels for the archive. These issues were also not found until the manual building of an archive release began. Often times an issue would be discovered right before the release of an archive data set and it would not be until a subsequent release for a resolution. Other times a science team would not review their data delivery from OPGS, thereby leading to a second look being missed and errors propagating to the next level where the PDS finds them and reports them back to OPGS. Sometimes issues could be fixed while the archive was being built but this would typically lead to an archive release being delayed. Fixing an archive issue while it was being built required more time from operators and made building the archive more expensive than anticipated. For later releases it would not be uncommon for data from past releases to be reprocessed and redelivered to resolve issues. Part of the reason for having issues early in a mission is that science teams often don't think about archiving early on when development is ongoing. It would not be until MSL archive release #10 that a data set was delivered to the PDS with no issues relating to the data or its labels. It was determined that automating as many aspects of the archive building process as possible would enable issues to be found earlier and thereby lower costs.

With the need to automate the archive building process and reduce costs to data providers the Multimission Ground System and Services(MGSS), part of the Advanced Multi-Mission Operations System(AMMOS) funded the AMMOS-PDS Pipeline Service(APPS) task. APPS developed a software pipeline to automate the creation of a data provider's archive. The pipeline is designed to attach to a mission's operational processing pipeline or run standalone. It transforms and validates products as they are generated so as to inform an operator of any issues as soon as they arise and allow for building the archive automatically on the fly. Experiences from MSL determined that this software would be useful for archive building. Had MSL used this software the overall quality of data sets delivered to the PDS would have been improved and the generation of

an archive made more seamless. It is recommended that data providers use software such as APPS's pipeline to automate archive building and thereby make archive building seamless and less costly.

APPS found other methods for decreasing costs to data providers. The PDS4 standard is now required for all future datasets delivered to the PDS. XML is used for PDS4 metadata labels and validation of PDS4 labels is done using XML schemas and schematrons. Prior to this designing a metadata label for a data product was simpler. PDS3 labels used simple key equals value entries for a metadata label. XML experience was not required. Not only must a data provider learn how to write XML for PDS4, they must also build a mission dictionary with mission specific attributes and classes. This mission dictionary is a set of schemas and schematrons that are specific to the mission or data set. PDS core and discipline dictionaries provide higher-level rules for the structure and content of a label that a mission dictionary's mission specific attributes and classes must abide by. This leads to PDS4 metadata labels having numerous rules that a data provider must learn in order to build a label. PDS3 had similar rules but they were not nearly as complex. PDS4's steep learning curve increases the cost and work required to design a metadata label for a PDS archive. Also, PDS4 requires labels for many more aspects of a PDS archive meaning not only is an individual label difficult to design, many different kinds of labels must be designed for each of the different files in the archive. PDS3 did not have this requirement.

APPS found that designing a PDS4 metadata label is difficult for non-experts. APPS designed the APPS Label Design Tool(LDT), a web-based drag and drop interface for building PDS4 labels, schemas, and schematrons without requiring any XML experience. Adding an attribute for a mission dictionary and a label is as simple as answering a few questions related to the type and description of the attribute. Adding a class is as simple as selecting what elements can be member of that class with checkboxes. Once an attribute or class is created in the LDT the user drags it from the mission dictionary pane to where the operator wants it to go in the label pane. Validation is done on the fly, allowing for instant feedback if the mission-specific attribute or class follows the rules specified by PDS4 core and discipline dictionaries. Once a label is built the user can generate a SIS Appendix, which uses the information provided in the mission dictionary that defines the label elements as well as the structure of the label to build an easy to read chart that defines all of the attributes and classes. This chart can then be imported directly into the SIS document. Updates to the mission dictionary automatically translate into this appendix once it is regenerated. For MSL there was sometimes a mismatch between label items in the SIS and label items in the label itself. By using the LDT this issue was removed and label creation became much more seamless.

InSight was APPS's first customer for its pipeline and Label Design Tool. The Mars 2020 rover has also expressed interest in using APPS tools and services. Both missions have determined that APPS can decrease the cost of building a PDS4 compliant archive. InSight has already used the Label Design Tool to assist with building its PDS4 labels and the APPS pipeline will be used to validate data products made from its Operational Readiness Test in February 2016. It is recommended that data providers consider APPS or similar tools and services to make their interaction with the PDS as seamless as possible.

Recommendations

I recommend that the PDS point of contact for data providers work as closely with the data provider as possible. In order to decrease costs and improve the quality of delivered data sets I recommend data providers use an automated pipeline for processing their data products. Specifically, I recommend using APPS's PDS pipeline. The data processing pipeline should be used as early as possible for missions so as to catch any issues with an archive as soon as they are apparent. Also, APPS tools should be used to aid data providers in building their metadata labels and SIS documents so as to decrease the workload for the PDS and learning curve for the data provider. In addition, by using APPS tools to build the SIS appendix for a data set, appendices will have a uniform structure across SIS documents.

Impacts

The impact of not making these suggested changes is that the interaction between PDS and data providers will remain inefficient and costly. Data providers will continue to face a steep learning curve to meet the PDS4 standard. This will force them to continue to rely on the PDS for assistance building their PDS4 archives thereby creating a higher burden for the PDS. Delivered data sets will not be as high of a quality as they could be. Data providers will continue to develop their own independent systems for data archiving and duplicate effort. Documentation costs will increase under PDS4 if integrated label development tools are unavailable.